

ESTIMASI POSE MODEL 3D DALAM LINGKUNGAN AUGMENTED REALITY BERBASIS TITIK FITUR WAJAH MENGGUNAKAN METODE POSIT

Heri Pratikno¹⁾

Mochamad Hariadi, ST., M.Sc., Ph.D²⁾

1) Bidang Keahlian Jaringan Cerdas Multimedia, Jurusan Teknik Elektro, Fakultas Teknologi Industri, Institut Teknologi Sepuluh Nopember Surabaya, email: heri@stikom.edu

2) Institut Teknologi Sepuluh Nopember, Surabaya 60111, email: mochar@ee.its.ac.id

Abstract: The main problem in objects tracking using cameras is to find the head pose estimation of the objects. In addition, it is necessary to rely on natural features since there is no light generated from the objects. This study aims to estimate the pose of a 3D head models using a single camera in a real-time *Augmented Reality* (AR) environment that is based on facial feature points. The position of the 3D models are translated into XYZ coordinate axes and rotated to the orientation angle RPY (*Roll, Pitch, Yaw*). The POSIT (*Pose from Orthography and Scale with Iteration*) is used for the pose estimation. The position and orientation of the 3D models are projected to facial feature points orthographically. In order to get the best pose, four to five iterations are performed in order to achieve the minimal error factor. The final results of this study can be considered as real-time systems as it can achieve 16 frames per second with an average angle rotational accuracy of 0.83° and translational deviation of 1.67 vector units.

Keywords: Augmented Reality, Facial Feature, Pose Estimation, POSIT.

Markerless dalam *Augmented Reality* (AR) bertujuan untuk menghasilkan interaksi antara komputer dan pengguna lebih alami dan intuitif dibandingkan dengan model *marker*. Deteksi yang berbasiskan titik fitur wajah merupakan salah satu model *markerless* yang banyak menarik perhatian para praktisi dan peneliti. Proses deteksi dan pengenalan yang berbasiskan titik fitur wajah dapat diterapkan dalam bidang keamanan, absensi, *e-commerce* dan *game*.

Proses estimasi pose (*pose estimation*) mempunyai tingkatan komputasi yang lebih sulit dan krusial dalam menentukan jumlah dan penempatan titik fitur wajah, transformasi model objek 3D serta akurasi, dibandingkan dengan proses deteksi wajah (*face detection*) maupun pengenalan wajah (*face recognition*).

Selain akurasi, proses *recovery fitting* antara *image* wajah 2D dan model objek 3D juga perlu diperhitungkan karena pergerakan terlalu cepat dari *image* wajah 2D aktor yang berfungsi sebagai *landmark* dari model objek 3D dapat menyebabkan kehilangan *frame* pelacakan dalam suatu *scene*.

Berdasarkan penelitian dari Daniel F. DeMenthon [DeMenthon, 1995], metode POSIT mempunyai komputasi sistem estimasi pose bisa

sampai dua puluh lima kali lebih cepat dari pada metode Yuan [Yuan, 1989] maupun metode Lowe [Lowe, 1985] karena pada metode POSIT sudah tidak diperlukan lagi proses inisialisasi estimasi pose awal yang dilakukan secara manual dan *inverse* matrik dalam *looping* iterasinya. Metode POSIT akan mengekstraksi pose objek dengan pendekatan model 3D secara antropometrik (*rigid anthropometric*) yang berkorespondensi dengan titik fitur *image* wajah.

Dengan dasar pemikiran tersebut maka penelitian ini membahas tentang implementasi "Estimasi pose model 3D dalam lingkungan *Augmented Reality* berbasis titik fitur wajah menggunakan metode POSIT". Beberapa metode sistem komputasi estimasi pose telah dikenalkan oleh Tsai [Tsai, 1987], Lowe dan Yuan.

Pada metode Lowe dan Yuan mempunyai dua kelemahan yang cukup signifikan, yaitu: pertama, sebuah perkiraan atau pendekatan pose harus dimulai dengan proses iterasi. Kelemahan kedua adalah setiap langkah proses iterasi diperlukan *pseudoinverse matrix* Jacobian dengan dimensi $2N \times 6$ pada metode Lowe dan $N \times 6$ untuk Yuan. Dimana N adalah jumlah titik fitur yang harus ditemukan, hal ini menjadikan sebuah operasi komputasional yang lebih lama dan rumit.

Metode *Fully Projective* [Araujo, 1998] merupakan pengembangan dari metode Lowe dan berhasil diimplementasikan dalam penelitian [Lim, 2002] dengan hasil penelitian sebagai berikut: tingkat kesalahan rotasi rata-rata sebesar 4° dan tingkat kesalahan translasinya rata-rata sebesar 5 cm.

Berdasarkan pembahasan pada latar belakang penelitian dan hasil penelitian sebelumnya, maka dirumuskan permasalahan pada penelitian ini, yaitu: proses interaksi dalam lingkungan *Augmented Reality* secara umum masih menggunakan *marker*, pada kasus *Augmented Reality* yang melibatkan interaksi dengan kepala manusia maka penggunaan *marker* kurang cocok untuk diimplementasikan.

Perlu adanya sistem yang mengimplementasikan metode POSIT agar proses interaksi dilingkungan *Augmented Reality* dapat dilakukan secara *markerless*, terkait dengan implementasi penggunaan metode POSIT tersebut maka perlu adanya pengujian terhadap performa estimasi pose pada POSIT dalam lingkungan *Augmented Reality*.

Penelitian ini bertujuan untuk memberikan metode alternatif dalam penerapan sistem interaksi antara manusia dan komputer dalam lingkungan *Augmented Reality* tanpa menggunakan *marker* (*markerless*). kemudian dilakukan uji performansi estimasi pose pada metode POSIT tersebut, yaitu: translasi pada posisi sumbu koordinat XYZ dan rotasi pada sudut RPY (*Roll, Pitch, yaw*).

Berdasarkan rumusan masalah, maka ada beberapa batasan masalah, antara lain :

- Jumlah orang pengendali gerakan model objek 3D pada lingkungan *Augmented Reality* hanya satu orang (aktor) dalam satu waktu.
- Pada penelitian ini tidak dibahas ekspresi dari model objek 3D.
- Kondisi pencahayaan merata dalam ruangan.

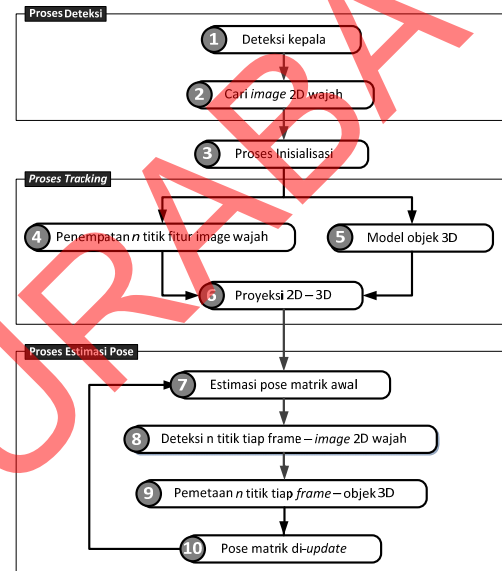
Hasil penelitian ini diharapkan memberi manfaat terwujudnya inetraksi di lingkungan *Augmented Reality* dapat dilakukan dalam waktu nyata secara *markerless*, yaitu: menggunakan titik-titik fitur wajah sebagai *landmark* penempatan (*layouting*) model objek 3D yang bisa digerakkan kearah enam taraf kebebasan bergerak (6DOF – *Six Degree of Freedom*). Adapun hasil dari penelitian ini dapat digunakan untuk industry game, absensi online, e-commerce, media promosi dan sebagainya.

Penelitian ini diharapkan bisa memberi kontribusi pada peningkatan interaksi yang lebih

alami dibandingkan dengan penggunaan *marker* serta lebih responsif dilingkungan *Augmented reality*, karena pada metode POSIT sudah tidak diperlukan lagi proses inialisasi pose awal titik-titik fitur di wajah yang dilakukan secara manual pada metode-metode sebelumnya.

METODE PENELITIAN

pada penelitian ini perancangan sistem secara garis besarnya dibagi menjadi tiga blok bagian utama, yaitu: proses deteksi, proses *tracking* dan proses estimasi pose. Sebagaimana tampak pada gambar 1.



Gambar 1. Bagan sistem

Proses Deteksi

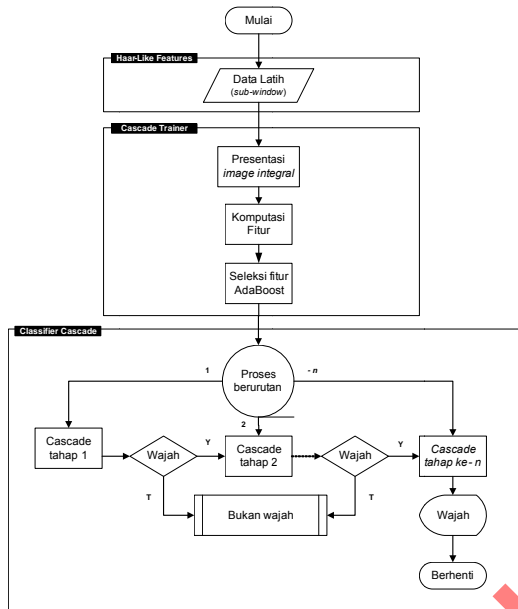
Pada tahapan ini bertujuan untuk pengambilan gambar (*grabbing*) tiap *frame* dalam waktu nyata (*real-time capturing*) dari sebuah *scene* menggunakan kamera tunggal.

Jarak ideal dari posisi objek *image* wajah dengan kamera adalah sekitar 15 cm (*near plane*) sampai dengan 3 meter (*far plane*). Proses *tracking* pada penelitian ini menggunakan metode Viola-Jones [Paul Viola, 2001], dimana fungsinya adalah untuk deteksi wajah (*face detection*) dengan tujuan untuk membedakan area wajah dan bukan wajah dari objek.

Algoritma Viola-Jones, terdiri dari tiga proses tahapan penting, yaitu: pertama adalah proses komputasi fitur (*feature computation*) yang dihasilkan oleh metode *Haar-Like Feature*. Proses kedua ialah seleksi fitur (*feature selection*) menggunakan metode pelatihan AdaBoost. Sedangkan proses yang ketiga adalah

ketepatan dalam waktu nyata (*real-timeliness*) dengan metode *cascade of classifier*.

Pada gambar 2 menunjukkan diagram alur tahapan proses pendeteksian, Hasil akhir dari proses deteksi *image* 2D kepala aktor pada penelitian ini, tampak pada gambar 3.



Gambar 2. Diagram alur proses deteksi metode Viola-Jones



Gambar 3. Posisi aktor terdeteksi

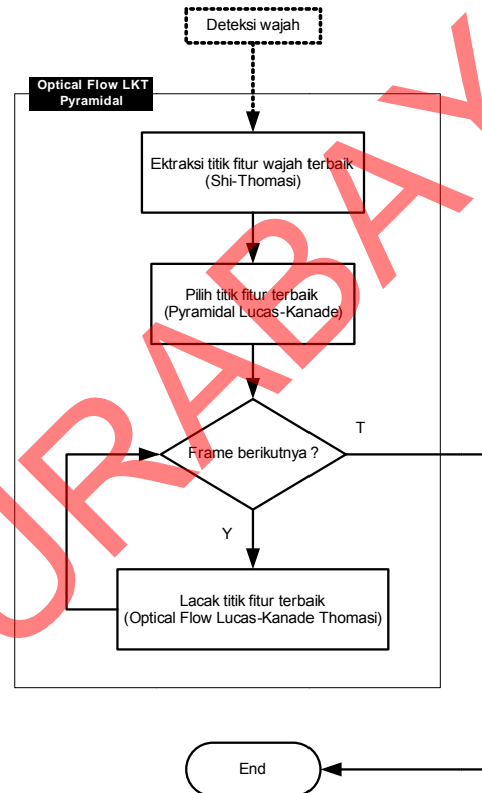
Tracking Wajah

Untuk melacak titik fitur wajah dari sebuah *image*, pada penelitian ini menggunakan algoritma Kanade-Lucas Thomasi (KLT) [Lucas and Kanade, 1981]. Pada prinsipnya metode KLT mempunyai tiga tahapan dasar dalam melacak titik fitur wajah, yaitu: mengekstraksi titik fitur wajah (*feature extraction*), memilih titik fitur wajah (*feature selection*) dan melacak titik fitur (*feature tracking*). Sebagaimana tampak pada gambar 4.

Pola tekstur hanya akan ada jika terlihat banyak piksel pada area tersebut, fitur yang terlacak akan lebih akurat apabila fitur *window* berisi informasi tekstur. Area pada fitur *window* bisa bermacam-macam tergantung dari

keperluan jumlah fitur yang dialokasikan, tampak pada gambar 5.

Optical flow adalah salah satu metode pelacakan objek secara *real-time*, dimana algoritma *optical flow* dapat diintegrasikan dengan algoritma yang lain untuk pelacakan dan pengenalan wajah.



Gambar 4. Proses Lucas-Kanade Thomasi



Gambar 5. ekstraksi fitur hasil penelitian



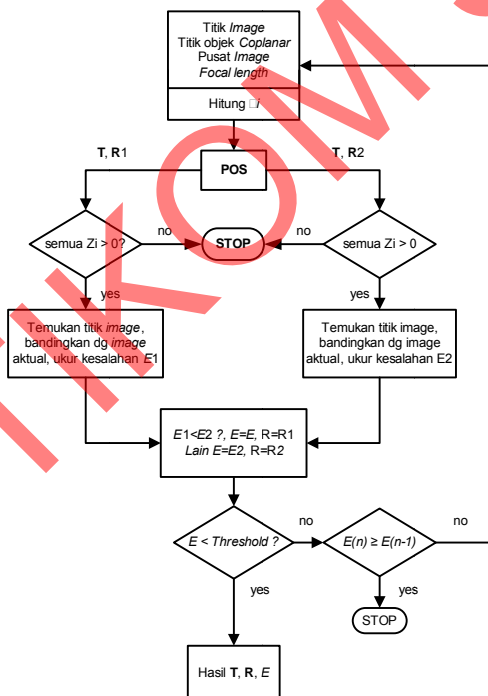
Gambar 6. Pelacakan *optical flow*

Estimasi pose

Pada dasarnya estimasi pose atau disebut juga *extrinsic camera calibration* adalah proses untuk mengekstraksi informasi yang terkait dengan informasi posisi dan orientasi titik yang korespondensi atau berkesesuaian dari sebuah model objek dan *image* yang diperoleh dari kamera. Pose dari sebuah objek merupakan posisi dan orientasi pada objek yang mempunyai enam arah kebebasan dalam bergerak (6DOF – *Six Degree of Freedom*).

Pada gambar 7, algoritma POS (*Pose from Orthography and Scale*) menghasilkan dua pose pada setiap iterasi pada algoritma POSIT. Akan tetapi dalam prakteknya hanya satu yang diikuti atau dua percabangan, yang akhirnya dengan satu atau dua solusi yang layak (+), kedua situasi tersebut terjadi karena :

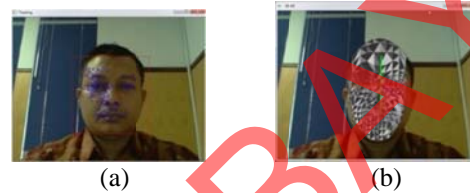
1. Pada situasi pertama, langkah iterasi pertama dikomputasi dengan dua pose tetapi satu pose tidak layak (-) atau dibuang karena beberapa titik-titik pose *scene* terletak dibelakang kamera. Sehingga hanya satu *path* layak (+) yang diproses.
2. Pada situasi kedua, kedua pose dari langkah iterasi pertama adalah layak (+), dan melanjutkan iterasi pada kedua cabang (*path*). Pada langkah kedua, masing-masing cabang masih tersedia dua pose yang layak (+), tetapi untuk masing-masing cabang hanya pose terbaik yang akan dipertahankan.



Gambar 7. Algoritma POSIT

HASIL DAN PEMBAHASAN

Implementasi sistem pada penelitian ini terdiri dari tiga bagian utama, yaitu: proses pertama adalah deteksi *image* 2D wajah, proses kedua adalah penentuan dan *tracking* titik fitur *image* 2D wajah tiap *frame*-nya, sedangkan proses ketiga adalah estimasi pose translasi kearah sumbu koordinat XYZ dan rotasi ke sudut RPY.



Gambar 8. Tampilan GUI window dari sistem, a. Window titik fitur wajah, b. Window Augmented Reality model objek 3D.

Pengujian Sistem

Untuk menguji sistem yang telah diimplementasikan pada penelitian ini akan dibahas dari sudut pandang ke-akurasian, ketangguhan pelacakan, pengaruh jumlah titik fitur *image* wajah, jumlah *frame* tiap detiknya serta estimasi pose.

ke-akurasian akan terkait dengan ketepatan dalam pelacakan, ketangguhan berhubungan dengan kemampuan melacak dalam perbedaan intensitas cahaya dan kecepatan gerakan. Banyaknya jumlah *frame* tiap detiknya akan memastikan proses komputasi dalam waktu nyata (*real-time*) atau tidak, sedangkan estimasi pose menunjukkan korelasi posisi dan orientasi antara model objek 3D dengan pergerakan atau pergeseran *image* 2D wajah aktor.

Akurasi Pelacakan

Pada implementasi sistem ini, akurasi pelacakan cukup baik karena bisa menggerakkan model objek kepala 3D pada layar animasi dengan pose aktor yang hampir sama dalam waktu nyata. Hasil pelacakan (*tracking*) mampu mendeteksi *image* wajah 2D dan menggerakkan model objek kepala 3D rotasi pada sudut *Pitch* ($X=0$) sebesar 129.1° ke arah $\pm Y$, rotasi pada sudut *Yaw* ($Y=0$) yang berpaling ke $\pm X$ besar simpangan sudut sebesar 179.8° . Sedangkan besarnya rotasi pada sudut *Roll* ($Z=0$) $\pm X$ adalah 180° .

Ketangguhan Pelacakan

Ketangguhan dalam pelacakan *image* wajah 2D serta pergerakan model objek kepala 3D dipengaruhi oleh pencahayaan ruangan dan pencahayaan pada aktor. Metode Lucas-Kanade Tomasi berdasarkan pada tingkat kecerahan objek (*brightness*) berjalan pada cahaya ruangan yang merata dan normal.

hal ini berarti pencahayaan pada ruangan maupun aktor tidak boleh terlalu gelap atau terlalu terang karena hal itu bisa mengakibatkan beberapa pelacakan titik fitur wajah bergeser dari tempat seharusnya. Apabila banyak titik fitur *image* wajah yang hilang atau bergeser maka bisa mengurangi tingkat akurasi pelacakan.

Estimasi Pose

Estimasi pose model objek 3D merupakan kombinasi dari rotasi matrik 3D dan translasi vektor 3D yang mempunyai nilai relatif dengan kamera, maka diperlukan minimal matrik 3 x 4. Karena pada OpenGL standar matriknya 4 x 4 (16 elemen) dengan prioritas perhitungan pada kolom (1D).

Jika menggunakan proyeksi perspektif maka hanya diperlukan empat titik *non-coplanar* (kiri, kanan, atas dan bawah) sedangkan pada proyeksi ortografi yang digunakan dalam OpenGL diperlukan enam buah titik *non-coplanar* (kiri, kanan, atas, bawah, dekat (*near*) dan jauh (*far*)).

Rotasi Pada Sumbu Z (Roll)



Gambar 9. Rotasi pada sumbu Z, a. di *window tracking*, b. di *window model 3D*.

Tabel 1. Rotasi pada sudut *Roll*

Frame ke-	Rotasi	Simpangan sudut	Deviasi		Rata-rata
	Di Sumbu Z	(derajat)	Antara tiap frame (°)	Semua frame (°)	
1	Roll (α)	14.703	0.482	1.271	0.623
2		15.185			
3		16.456			
4		16.859	0.403	0.99	
5		17.849			
6		19.175			
7		18.965	1.326	0.21	
8		20.160			
9		20.962			
10		21.615	1.195	0.802	
11		22.189			
12		22.457			
13		23.262	0.268	0.574	
14		23.951			
15		24.412			
16		24.593	0.6889	0.805	
17		24.91			
18		25.47			
19		25.441	0.181	0.461	
		0.56	0.317		
		0.029	0.029		

Rotasi Pada Sumbu Y (Yaw)

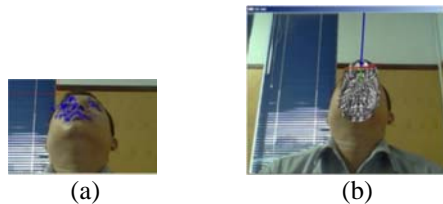


Gambar 10. Rotasi pada sumbu Y, a. di *window tracking* b. di *window model 3D*

Tabel 2. Rotasi pada sudut *Yaw*

Frame ke-	Rotasi Di Sumbu Y	Simpangan sudut (derajat)	Deviasi		Rata-rata
			Antara tiap frame (°)	Semua frame (°)	
1	Yaw (β)	72.668	0.828	1.006	1.007
2		73.496			
3		74.502			
4		76.162	1.66	1.778	
5		77.94			
6		78.621			
7		79.361	0.681	0.74	
8		79.363			
9		78.204			
10		79.207	0.002	1.159	
11		79.199			
12		79.59			
13		79.982	1.003	0.008	
14		80.308			
15		79.891			
16		79.235	0.391	0.392	
17		74.611			
18		76.986			
19		76.912	2.375	4.624	
		0.074	0.074		

Rotasi Pada Sumbu X (Pitch)



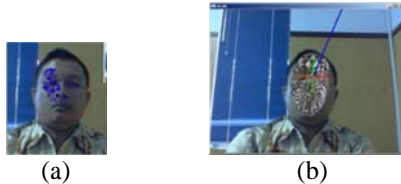
Gambar 11. Rotasi pada sumbu Z, a. di *window tracking*, b. di *window model 3D*.

Tabel 3. Rotasi pada sudut *Pitch*

Frame	Rotasi	Simpangan sudut	Deviasi		Rata-rata
ke-	Di Sumbu X	(derajat)	Antara tiap frame (°)		Semua frame (°)
1	Pitch (β)	-25.086	0.206	1.726	0.869
2		-25.292			
3		-23.566	1.168	0.193	
4		-22.398			
5		-22.591	3.865	0.593	
6		-26.456			
7		-27.049	0.052	2.003	
8		27.101			
9		-25.098	0.925	0.2	
10		-24.173			
11		-24.373	1.143	0.21	
12		-25.516			
13		-25.726	0.664	0.126	
14		-26.39			
15		-26.516	0.883	1.26	
16		-27.399			
17		-28.659	0.418	0.001	
18		-29.077			
19		-29.078			

Translasi Pada Sumbu XYZ

Translasi dilakukan pada ketiga sumbu XYZ, karena keterbatasan jumlah halaman, maka hanya translasi kearah sumbu Z yang digunakan sebagai contoh.



Gambar 4.25 Translasi dari +Z ke -Z, a. di *window tracking*, b. di *window Augmented Reality* model 3D.

Tabel 4. Translasi ke-15 langkah pada sumbu Z

Translasi Ke-	Sumbu	Arah	Nilai	Rata-rata	Satuan
1	Z	Depan ke Belakang	1.4425	1.7344	Unit Vektor
2			1.453		
3			1.4696		
4			1.5157		
5			1.5528		
6			1.6083		
7			1.6653		
8			1.7201		
9			1.8087		
10			1.8608		
11			1.8838		
12			1.9347		
13			1.9955		
14			2.0266		
15			2.0841		

KESIMPULAN

Dari hasil perancangan sistem kemudian dilanjutkan pengambilan data, pengujian dan analisa, maka dapat disimpulkan bahwa pada penelitian ini telah berhasil diimplementasikan dengan hasil sebagai berikut:

1. Metode POSIT dapat digunakan untuk estimasi pose pada lingkungan *Augmented Reality* dengan hasil interaksi lebih alami karena sudah tidak menggunakan *marker* lagi tetapi berbasis titik fitur wajah.
2. Hasil keluaran dari sistem yang diterapkan

sebesar 16 *frame per second* (FPS), sehingga sudah termasuk dalam waktu nyata (*real-time*) yang men-syarat-kan minimal 4 FPS.

3. Diperlukan 200 titik fitur pada *image* wajah aktor sebagai *landmark* model objek 3D agar tidak bergetar (*tremor*).
4. Estimasi pose dicapai dengan rata-rata deviasi orientasi sudut 0.83° dengan rata-rata deviasi jarak translasi 1.67 satuan unit vektor.

Berikut adalah beberapa saran untuk penelitian lebih lanjut:

1. Penerapan metode Fuzzy, AMM dan sebagainya untuk mendeteksi ekspresi wajah aktor.
2. Untuk mengatasi efek *Gimbal Lock* direkomendasikan menggunakan metode Quaternion (4D).

RUJUKAN

- Araujo H., Carceroni Rodrigo L. Brown Christopher M., *A Fully Projective Formulation to Improve the Accuracy of Lowe's Pose Estimation Algorithm*, Computer Vision and Images Understanding, Volume 70, Pages 227-238, 1998.
- B. D. Lucas and T. Kanade., *An iterative image registration technique with an application to stereo vision*, Proceedings of the DARPA imaging understanding workshop, pp. 121-130. 1981.
- Daniel F. DeMenthon, Larry S. Davis., *Model-Based Object Pose in 25 Lines of Code*. International Computer Vision Laboratory, University of Maryland, 1995.
- Lowe, D.G., *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, 1985.
- Paul Viola, Michael Jones., *Robust Real-Time Object Detection*. Canada. July 13, 2001.
- Resmana Lim. Davina. Sivia R., *Pelacakan dan estimasi Pose Video wajah 3 Dimensi*, Jurnal Teknik Elektro, volume 2, nomer 2, UK Petra, 2002.
- R.Y. Tsai., *An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision*. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, pp. 364-374, 1986.
- Yuan, J.S.C., *A General Photogrammetric Method for Determining Object Position and Orientation*, IEEE Trans. on Robotics and Automation, vol.5, pp.129-142, 1989.